

Review Comments

The research article addresses the issue of instability of LDA topic inference due to random initialization of topic. They proposed a novel approach, named LDAPrototype, to reproduced the same results across multiple runs of topic inference. They observed that LDAPrototype resulted remarkable similarity between replication of topic inference. They proposed a similarity measure S-CLOP to examine the stability of LDA. By the experiments, they proved that S-CLOP outperformed the measures of quality assessment like NPMI and perplexity. They also suggested robust version of the Jaccard coefficient for topic evaluation.

Basic Reporting

- The entire manuscript has been produced by the authors in professional, precise English. Their attention to punctuation, spelling and grammars appears to have been significant during the manuscript preparation process. The writing style is professional, clear, and concise.
- The related work section demonstrates a thorough understanding of the existing methodology. The review successfully identifies specific gaps in the existing literature. However, authors could have included some relevant work, if any, by other researchers.
- Figures and tables are appropriately captioned and labelled, relevant and well-constructed. Figures quality have been maintained across the paper.

Experimental design

- The experimental work is designed and executed with a high degree of methodological rigor. The research provides detailed protocols for the experimental procedures, including precise measurements.
- Datasets have been explained and presented well. However, I would suggest dataset description should be presented in tabular format.

Validity of the findings

- Experimental outcomes have been validated by variety of runs. The experiments have been performed well across the parameters settings. Also, different datasets have been considered for the numerous experiments. Overall, the experimental works have been carried out in 360 degree view.

General comments

- Authors are suggested to explain ‘the reliability of topic model’ in more detail. Give some appropriate example for better understanding of the article.
- The S-CLOP is an abbreviated term. Please mention full-form of it in the introduction section. Therefore, reader can interpret it better.
- Authors should cite recent research articles in the domain under study. They should discuss the limitations and issues of the recent papers.
- Your related work needs more detail. I suggest that you include similar research work.
- Authors should have explained the term ‘replication’ with respect to LDA topic inference for better understanding of the readers.
- I suggest authors should explain how the medoid is chosen.
- Page No. 6 – It seems that first line is not part of text. Please check and remove if not needed. I suggest to remove if it is not needed.
- Line No. 224 – “Then the deviation from the perfect situation of one representative from each LDA run is quantified.” Please explain in detail how it is done.
- Line No. 225, authors have mentioned ‘representative’ of each LDA run, Please elaborate this terminology.
- Line No. 237, authors have explained LDA disparity. I suggest, if possible, authors to take an example for the readers. Why is disparity calculated? Please describe.
- Line No. 248 and 249 needs to be revised, as it seems less interpretable.
- As you have removed the words with total count less than some limit (5 in this study) to reduce vocabulary size, have you considered upper bounds to remove the words that appear very frequently?
- Relative limit is 0.002, and it results in 100 words per topic as relevant. Generally, top 10 or 20 words are chosen for the topic representation. Could you please explain the reason behind considering 100 words per topic?